Harnessing the power of longitudinal data

David Barber MD Rachael Morkem MSc John A. Queenan PhD Karen Hall Barber MSc(HQ) MD

The Canadian Primary Care Sentinel Surveillance Network (CPCSSN), a health informatics database, now contains the electronic medical chart data of more than 1 million patients provided by 1000 sentinel physicians using 12 different electronic medical record (EMR) products. The purpose of the CPCSSN database is to provide a national platform for surveillance, research, evaluation, and practice quality improvement. For example, CPCSSN can identify patients with diabetes by analyzing ICD-9 codes, medication lists, and laboratory test results to create a validated case definition for diabetes that can be used to report national prevalence estimates.[1,2] CPCSSN creates synoptic chronic disease registries using information contained within practice-based EMRs by cleaning, coding, and calculating derived variables, which are shared with participating physicians and clinics. The creation of a disease registry is the foundation of quality improvement within a practice and, in this vein, CPCSSN is invaluable to practices that have low "data discipline."[3]

CPCSSN is also capable of analyzing longitudinal data, some of which were collected before 2006, in order to highlight changes over time. Longitudinal data can provide information that can be used to predict clinical outcomes and create clinical decision support tools (predictive analytics). Predictive analytics, a hot topic in the health information technology world, use data mining techniques, such as machine learning, statistical analysis, and modeling, to determine patterns and subtle relationships within data and to infer rules that will reveal complex and hidden relationships in large amounts of data.[4]

For example, would it not be helpful to be able to tell patients their 5-year risk of developing diabetes, as is done with cardiovascular risk? Using longitudinal data, CPCSSN researchers are able to identify variables associated with the development of diabetes, such as rising hemoglobin A_{1c} levels that are not yet abnormal, weight changes, and prescribing patterns. This in turn could lead to the development of computerized alerts or reminders in the corner of an EMR that report a patient's 5-year risk of developing diabetes. This value might promote dialogue between patients and their health care providers and help to motivate patients to make lifestyle changes that might reduce the risk of developing diabetes.

Another example of the power of longitudinal data was noted in one of our practices. An elderly patient developed colon cancer and on review of his chart data, it was noted that his mean corpuscular volume (MCV) had been dropping over a 2-year period, suggesting ongoing blood loss from his cancer. Should a dropping MCV in an elderly person trigger a warning to the health care provider? In the past, without straightforward access to primary care data, this question would have been difficult to answer. Now, with the access to such data and the ability to link to provincial databases, this question can and should be answered. If a downward trend of the MCV in an elderly person is substantial, then a decision support tool can be developed within EMRs that can appropriately alert the health care provider.

These "bench-to-bedside" scenarios illustrate the ways longitudinal primary care data can power predictive analytics, resulting in real-world interventions in the form of clinical decision support tools. Building predictive models using a vast database like CPCSSN will lead to a better understanding of patient populations and allow primary care providers to identify and mitigate risks within these populations.

Dr Barber is Assistant Professor in the Department of Family Medicine at Queen's University in Kingston, Ont. **Ms Morkem** is Research Associate and Privacy Officer for CPCSSN. **Dr Queenan** is an epidemiologist at the Centre for Studies in Primary Care at Queen's University and with CPCSSN. **Dr Hall Barber** is Assistant Professor in the Department of Family Medicine at Queen's University.

Acknowledgment

Funding for this publication was provided by the Public Health Agency of Canada. The views expressed do not necessarily represent the views of the Public Health Agency of Canada.

Competing interests

None declared

References

1. Williamson T, Green ME, Birtwhistle R, Khan S, Garies S, Wong ST, et al. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. *Ann Fam Med* 2014;12(4):367-72.
2. Greiver M, Williamson T, Barber D, Birtwhistle R, Aliarzadeh B, Khan S, et al. Prevalence and epidemiology of diabetes in Canadian primary care practices: a report from the Canadian Primary Care Sentinel Surveillance Network. *Can J Diabetes* 2014;38(3):179-85. Epub 2014 May 14.
3. Han W, Sharman R, Heider A, Maloney N, Yang M, Singh R. Impact of electronic diabetes registry 'Meaningful Use' on quality of care and hospital utilization. *J Am Med Inform Assoc* 2015 Jul 1. Epub ahead of print.
4. Lumin S, Lu J. *Healthcare data manipulation and analytics using SAS.* Paper PO17. Paper presented at: PharmaSUG 2014; 2014 Jun 1-4; San Diego, CA.

Sentinel Eye is coordinated by CPCSSN, in partnership with the CFPC, to highlight surveillance and research initiatives related to chronic illness prevalence and management in Canada. Please send questions or comments to Dr Richard Birtwhistle, Chair, CPCSSN, at richard.birtwhistle@dfm.queensu.ca.

Cet article se trouve aussi en français à la page 356.